

Netflix Siblings

Taylor Blair

1/26/2021

#A brief intro

My sister and I have a long history of watching Netflix.

There are some minor outliers in the datasets, namely when we use each others accounts.

#Loading data and libraries

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.3      v purrr  0.3.4
```

```
## v tibble  3.0.5      v dplyr  1.0.3
```

```
## v tidyr   1.1.2      v stringr 1.4.0
```

```
## v readr   1.4.0      v forcats 0.5.0
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x dplyr::filter() masks stats::filter()
```

```
## x dplyr::lag()     masks stats::lag()
```

```
library(openair)
```

```
library(readr)
```

```
my_history <- read_csv("MyNetflixHistory.csv")
```

```
##
```

```
## -- Column specification -----
```

```
## cols(
```

```
##   Title = col_character(),
```

```
##   Date = col_character()
```

```
## )
```

```
sibling_history <- read_csv("SisNetflixHistory.csv")
```

```
##
```

```
## -- Column specification -----
```

```
## cols(
```

```
##   Title = col_character(),
```

```
##   Date = col_character()
```

```
## )
```

#Data prep

##Date fixing

```
my_history$Date <- as.Date(my_history$Date, "%m/%d/%y")
```

```
sibling_history$Date <- as.Date(sibling_history$Date, "%m/%d/%y")
```

```
##Movie or TV series
```

```
my_history$type <- grepl(":", my_history$Title)
my_history$type <- my_history$type %>%
  replace(my_history$type==TRUE, "TV Series") %>%
  replace(my_history$type==FALSE, "Movie")

sibling_history$type <- grepl(":", sibling_history$Title)
sibling_history$type <- sibling_history$type %>%
  replace(sibling_history$type==TRUE, "TV Series") %>%
  replace(sibling_history$type==FALSE, "Movie")
```

```
##Series, Season and episode
```

```
my_history <- my_history %>%
  separate(Title,
    c("Series", "Season", "Episode"),
    ": ")
```

```
## Warning: Expected 3 pieces. Additional pieces discarded in 267 rows [23, 99,
## 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115,
## 116, 117, ...].
```

```
## Warning: Expected 3 pieces. Missing pieces filled with `NA` in 40 rows [5, 50,
## 56, 87, 88, 151, 152, 153, 291, 300, 301, 306, 308, 337, 379, 501, 561, 656,
## 806, 901, ...].
```

```
sibling_history <- sibling_history %>%
  separate(Title,
    c("Series", "Season", "Episode"),
    ": ")
```

```
## Warning: Expected 3 pieces. Additional pieces discarded in 263 rows [105, 106,
## 107, 108, 115, 116, 127, 128, 138, 139, 355, 399, 400, 401, 402, 403, 404, 405,
## 406, 407, ...].
```

```
## Warning: Expected 3 pieces. Missing pieces filled with `NA` in 53 rows [1, 110,
## 174, 175, 261, 262, 263, 275, 276, 281, 282, 283, 284, 285, 454, 460, 515, 516,
## 521, 703, ...].
```

```
#Day of Week
```

```
day_of_week <- function(x){
  return (weekdays(x))
}
```

```
my_history$day <- sapply(my_history$Date, day_of_week)
sibling_history$day <- sapply(sibling_history$Date, day_of_week)
```

```
#Compare dataset
```

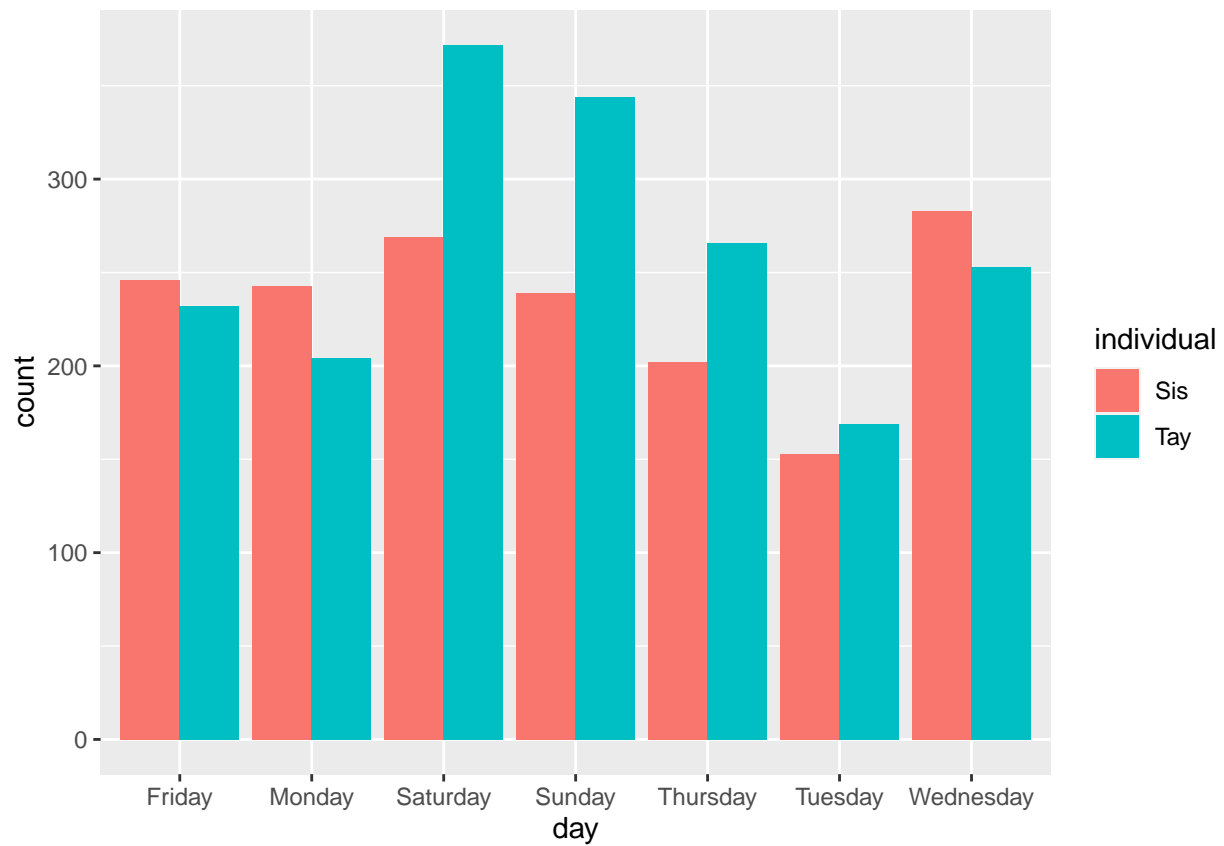
```
my_history$individual <- "Tay"
sibling_history$individual <- "Sis"
```

```
merged <- rbind(my_history, sibling_history)
```

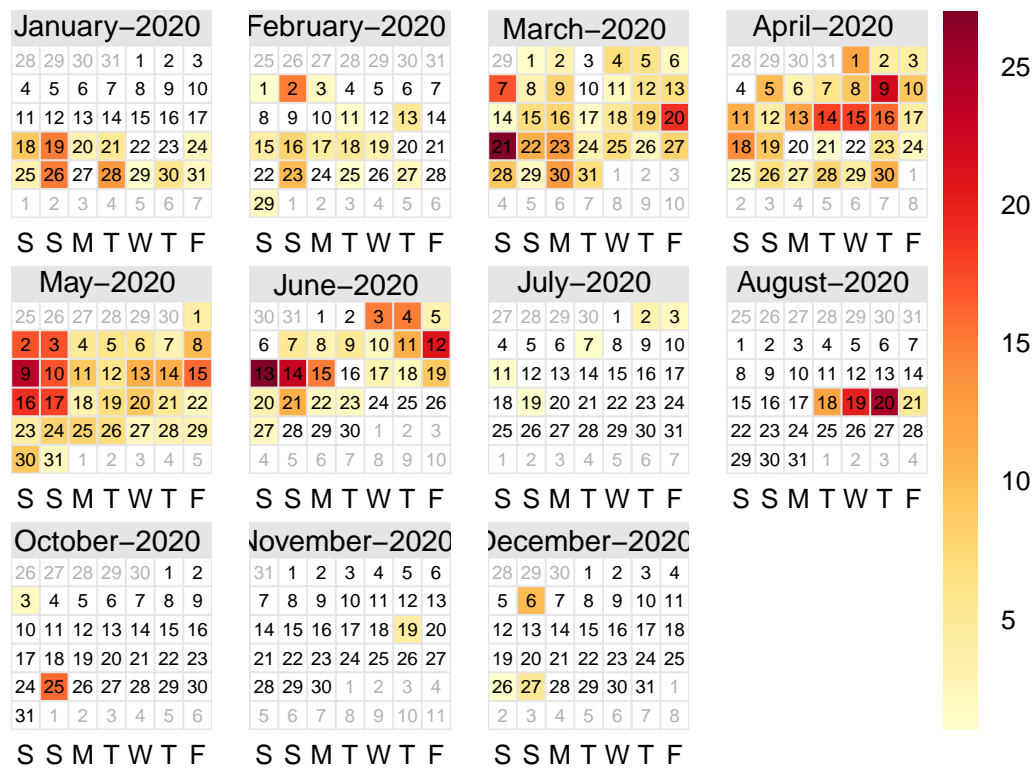
```
#Numerical analysis
```

```
#Graphical Analysis
```

```
ggplot(merged, aes(x = day, fill=individual) ) +
  geom_bar(position = "dodge") +
  labs("Sis v. Tay: Netflix shows watched by day")
```



```
my_history %>%
  group_by(Date) %>%
  summarise(count = n()) %>%
  rename(date=Date) %>%
  calendarPlot(pollutant = "count",
    year = 2020)
```



```
sibling_history %>%
  group_by(Date) %>%
  summarise(count = n()) %>%
  rename(date=Date) %>%
  calendarPlot(pollutant = "count",
               year = 2020,
               month = c(1:12))
```

